

Data Sharing

Moderator:

David Baldwin
Ecological Society of America
Journals
Ithaca, New York

Speakers:

Michael Whitlock
The American Naturalist
Vancouver, British Columbia

Francis Ouellette
Ontario Institute for Cancer Research
Toronto, Ontario

Reporter:

Olga Kuchment
University of California
Berkeley, California

Charles Elton, a highly influential ecologist, amassed decades of data on the woods of England—invaluable data for future ecological research. Unfortunately, the data disappeared after Elton's death in 1991. This story was told by David Baldwin, managing editor of the Ecological Society of America (ESA) journals, to illustrate why public archiving of data is needed. Baldwin described ESA's efforts to encourage data archiving: publishing archive journals, offering to publish abstracts of archived data in *Ecology*, and requiring that data used in publications be shared. The response from the scientific community has been weak.

Michael Whitlock, editor-in-chief of *The American Naturalist*, and Francis Ouellette, associate director of informatics and bio-computing at the Ontario Institute for Cancer Research, spoke of the difficulties in scientific-data archiving and ways to overcome them. Whitlock has coordinated a movement in ecology and evolutionary-biology journals to require public archiving of data. Ouellette, formerly coordinator of the open database GenBank at the

National Institutes of Health, has many years of experience with public archiving in molecular biology, in which public archiving is now the norm.

Public archiving of data is useful for many reasons, Whitlock said. One reason is error checking: Several studies show that 5% to 10% of published papers have statistical errors that change the conclusions. Other major reasons are to enable the meta-analysis of numerous sets of data or entirely new uses of the data. Another study showed that articles whose data are publicly archived are cited more often.

Throughout his career, Ouellette has observed the rapid increase in publicly archived molecular-biology data. Funding agencies and publishers helped institute and promote policies requiring data archiving. The policies are updated often to deal with new data types and to place new data into the public domain as quickly as possible. Since the early 1990s, all new DNA-sequence data must be in GenBank once they are published.

Whitlock said that most journals in ecology and evolutionary biology require data sharing, but not archiving, for publication. Data sharing is hampered by disorganization and lack of clear policies, so data continue to be lost.

Although public data archiving is a way to enforce data sharing, there are obstacles. The diverse data of evolutionary biology and ecology are technologically difficult to archive; this difficulty is being addressed. More difficult to overcome is the cultural inertia against archiving. Archiving takes time and effort: To be useful, preserved data must be organized and accompanied by metadata, such as their origin and a list of contents. In addition, authors often wish to keep data hidden from possible competitors.

Journals and archives can overcome the obstacles by working together, Whitlock said. Once it is technologically feasible,

the most prominent journals in evolutionary biology will require public archiving of data as a condition for publication; ecology journals may do the same. The process of archiving will be streamlined so that the only metadata required will be the authors' names and contact information and the article citation. Authors will be able to keep data sets from becoming public for a short time.

Ouellette responded that using minimal metadata will be satisfactory only if both the papers and data are publicly available. He said that funding agencies and academic institutions, not journals, should enforce data archiving by making archiving a condition for funding and promotion. Journals should support the policies, and users of the data should notify archives of errors. If those steps are taken, Ouellette said, the cultures of scientific fields will slowly change until public archiving is the norm. 🔗